A symbolic method to analyse patterns in plant structure whose organogenesis is driven by a multitype branching process

Cedric Loi, Paul-Henry Cournède, Jean Françon

A symbolic method to analyse patterns in plant structure whose organogenesis is driven by a multitype branching process

Cédric Loi[1,2], Paul-Henry Cournède[1,2], Jean Françon[3]

[1] *Ecole Centrale Paris, Laboratory of Applied Mathematics and Systems, 92290 Châtenay Malabry, France*
[2] *INRIA Saclay - Île-de-France, DIGIPLANTE, 91400 Orsay, France*
[3] *Université de Strasbourg, 67000 Strasbourg, France*

E-mail: cedric.loi@ecp.fr paul-henry.cournede@ecp.fr jean_francon@hotmail.fr

Received March the 23^{th}

Abstract Formal grammars like L-systems have long been used to describe plant growth dynamics. In this article, they are used for a new purpose. The aim is to build a symbolic method derived from computer science that enables the computation of the distribution associated to the number of complex structures in plants whose organogenesis is driven by a multitype branching process. To that purpose, a new combinatorial framework is set in which plant structure is coded by a Dyck word. Moreover, the organogenesis is represented by stochastic F0L-systems. By doing so, the problem is equivalent to determining the distribution of patterns in random words generated by stochastic F0L-system. This method leads directly to numerous applications like parametric identification for plant growth model.

Keywords Dyck word, stochastic F0L-system, plant organogenesis model, symbolic method

1 Introduction

In botany, the organogenesis is the process during which the buds of a plant produce new organs. Formal grammars have long been used to describe plant organogenesis. In particular, the parallel rewriting grammar introduced by [1] (called L-system) is well adapted to model the evolution of branching patterns and its algorithmic power has been broadly taken advantage of since [2]. It has thus provided efficient algorithms and subsequently software language for plant simulation (see [3]). For structures as complex as trees, bud production rules are influenced by many factors and are usually modelled by stochastic processes. In that case, the organogenesis is represented by stochastic 0L-systems. This type of grammar gives interesting results from simulation and graphical points of view by increasing the realistic aspect of geometric plants (see [4]). However, their full mathematical potential has not been taken advantage of.

In this article, stochastic 0L-systems are used for a new purpose. The objective is to write a method allowing the computation of the distribution associated to the number of complex structures in plants whose organogenesis is driven by a multitype Galton-Watson branching process. This method relies on a symbolic

approach derived from theoretical computer science and the analysis of algorithm (see [5]). Plants are seen as combinatorial structures. In this new combinatotial framework, plant topology is described by a Dyck word. The evolution of the structure of the plant is given by a set of stochastic rules contained in a stochastic F0L-system. Therefore, this framework allows the use of powerful methods of combinatorics such as the symbolic method (see [5]). Including a symbolic approach in a dynamic branching structure gives a more complete description of the system. It enables the computation of the distribution of patterns in a sequence of words whose dynamic evolution is driven by a branching process. Such results have many applications: comparison of stochastic models, parameter identification ...

Some basic concepts of botany and the main features of stochastic organogenesis models driven by a multitype branching process are first recalled in Section 2. Then, a new combinatorial framework is set up in Section 3. Plants are seen as labelled plane rooted trees. It allows the description of their structures by Dyck words. The evolution rules of the organogenesis model are represented by a stochastic F0L-system. Section 4 proposes a symbolic method adapted to plant growth models. In the last section, we show how to use the symbolic method for the parameter identification of the stochastic processes during plant development.

2 Stochastic modelling of organogenesis with a multitype branching process

Models of plant development (or organogenesis) describe the dynamic creation of organs (internodes, buds, leaves, flowers or fruits) and how they arrange to form plant structure. When the smallest scale of interest is that of organs (and not cells), discrete models are generally used to simulate plant structural development.

2.1 Modelling of plant structure

In this article, only the above ground parts of plants are considered. As explained in [6], organogenesis results from the functioning of undifferentiated cells constituting the apical meristem and located at the tip of axes. When in active phase, this meristem forms buds that will develop into agglomerates of organs composed of one or several ***phytomers*** (also called metamers). A phytomer is a botanical entity chosen as the elementary unit used to represent the plant architecture. It is composed of an internode bearing buds (terminal and lateral) and a leaf (see Figure 1). Depending on the type of plant, the internode may also bear flowers and fruits.

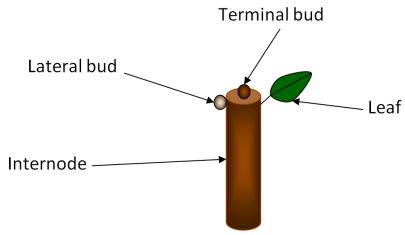

Fig. 1. Example of phytomer

Concerning the architecture of the plant, axis and architectural units can be listed into different categories depending on their morphological parameters (length, diameter, ...). These categories will be called ***Morphological Categories*** (= MC) in the sequel. Other names can be found in the literature (for example, [6] introduced the concept of physiological age to represent the different types of axes resulting from the meristematic differentiation in a plant; the concept of MC is more general and is chosen to allow cases that do not correspond to the strict botanical concept of physiological age). By convention, the terminal bud of an axis is thus characterized by the MC of the phytomer that bears it (Figure 2 gives an example of plant with two MCs).

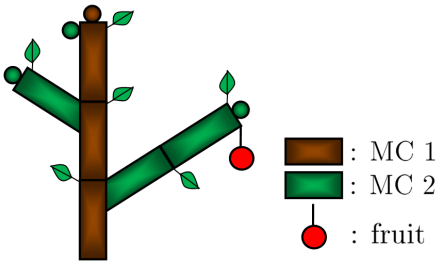

Fig. 2. Example of plant with two Morphological Categories (= MC). The fruit has no MC attached.

In the sequel, an organ will be characterized by its type (internode, bud, fruit, ...) and by its botanical characteristics (MC, Chronological Age, ...).

2.2 Stochastic organogenesis

Plant development can be discretised in time. The time between the appearances of new shoots (*i.e.* the time step of the discrete model) defines the ***Growth Cycle*** (= GC). For example, most temperate trees grow rhythmically, new shoots appearing at spring. If we do not consider polycyclism and neoformation, the GC corresponds to one year. The ***Chronological Age*** (= CA) of a plant (or of an organ) is defined as the number of GCs it has existed for.

The structure of a plant changes from one GC to another. For example, a bud may create new organs or the botanical characteristics of an organ may change (such as its CA). A set of rules called ***evolution rules*** defines the way organs evolve from one GC to another. In a stochastic organogenesis model, an organ may have several possible evolutions. In that case, a probability of occurrence is associated to each of these. As explained in [7], the probabilities involved in the organogenesis model are the result of botanical phenomena (bud survival, bud dormancy, differentiation, ...).

In this article, we consider a class of organogenesis models that satisfy the following hypotheses:

1. at each GC, the organs behave independently: the evolution of an organ is not influenced by the other organs of the plant.

2. the probabilities of evolution associated to an organ depend only on its type and its characteristics.

Under these hypotheses, a plant structure of CA N is built recursively by using the following algorithmic procedure:

- Initialisation: the structure of the plant at GC 0 is given by a seed (*i.e.* a bud of MC 1).

- for all $n = 0 \cdots N-1$: the plant structure at GC $n+1$ is built from the plant structure at GC n by replacing randomly all the organs by one of their possible evolutions according to their probabilities of occurrence.

The underlying stochastic process associated to this class of organogenesis models is a multitype Galton-Watson branching process (see [8] and [9]). As a matter of fact, the organs used to build plant structure can be seen as individuals in a population process. Therefore, the previous algorithmic procedure creates a Markovian sequence of random plants indexed by the GCs. The evolution rules associated to the organogenesis model are usually represented by a set of stochastic automata (see [2] and [4] and Figure 3).

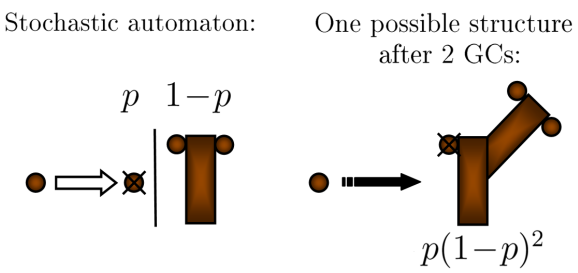

Fig. 3. Example of stochastic automaton for an organogenesis model

Figure 3 shows an example of organogenesis model with one MC. p represents the death probability for a bud. In this example, a bud has a probability p to die and a probability $1-p$ to create one internode with two lateral buds. Note that there is no stochastic automaton concerning the possible evolutions of an internode. In that case, it means that internodes do not evolve from one GC to another (they stay in the same state with a probability equal to 1). The plant on the right is one possible structure occurrence after two GCs. The associated probability of occurrence is $p(1-p)^2$.

In the sequel, $\mathcal{S}$ will denote a stochastic organogenesis model driven by a multitype branching process. A 'plant' refers to a branched structure generated from a seed (*i.e.* a bud of MC 1) according to the evolution rules of $\mathcal{S}$. In the following section, a combinatorial framework is set up to describe plant architecture and its evolution. The structure of a plant is given by a Dyck word in a bijective way. Since the organogenesis is driven by a multitype Galton-Watson branching process, the evolution rules can be represented by a stochastic 0L-system (see [10] for more details).

3 Plants as combinatorial structures

3.1 Some combinatorial concepts

We recall some basic definitions and properties of combinatorics (see [5] and [11] for more details).

Definition 1 (Plane Rooted Tree = PRT). *A rooted tree is a connected and acyclic graph with a node specifically distinguished called root. A plane tree is defined as a tree in which subtrees coming from a common node are ordered and represented from left to right.*

Figure 4 gives examples of PRT. The trees (2) and (3) are equivalent as rooted trees, but they become distinct objects when regarded as PRTs.

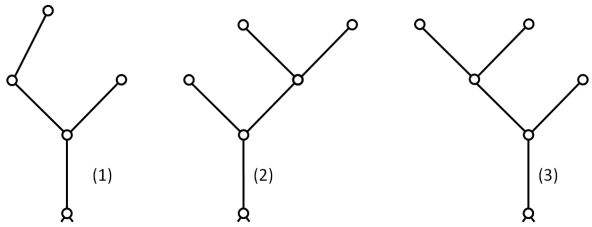

Fig. 4. Examples of plane rooted trees

PRTs can be coded in several ways. One of the most classical one is the Dyck word coding (see [12]). The latter relies on a tree traversal using the prefix order (see [5], page 74):

Definition 2 (Prefix order). *A PRT is traversed according to the prefix order if it is traversed starting from the root, proceeding depth-first and left-to-right, and backtracking upwards once a subtree has been completely traversed.*

Figure 5 gives an example of preorder traversal.

Definition 3 (Dyck word coding). *Every PRT is described by a Dyck word on the alphabet $V_1 = \{z, z'\}$ as follows:*

- *the tree is traversed according to the prefix order.*
- *an edge visited from the parent node to the child node is represented by the letter z.*
- *an edge visited from the child node to the parent node is represented by the letter z'.*

In the case of labelled trees (trees with characteristics attached to nodes and edges), a

new code deriving from Dyck word coding can be used:

Definition 4 (Extended Dyck word coding). *Let $L = \{l_1, \ldots, l_M\}$ be a set of labels. Every labelled PRT with labels in L is described by an extended Dyck word on the alphabet $V_2 = \{z_{l_1}, z'_{l_1}, \ldots, z_{l_M}, z'_{l_M}\}$ as follows:*

- *the tree is traversed according to the prefix order.*
- *an edge with a label l_i and visited from the parent node to the child node is represented by the letter z_{l_i}.*
- *an edge with a label l_i and visited from the child node to the parent node is represented by the letter z'_{l_i}.*

3.2 Coding a plant structure with a Dyck word

Every plant generated by the organogenesis model $\mathcal{S}$ can be represented by a labelled PRT:

Definition 5 (Labelled PRT associated to a plant). *Let P be a plant generated by the organogenesis model $\mathcal{S}$. Let us build a labelled PRT from the plant P as follows:*

- *Every organ is represented by an edge and ends with a node at each extremity.*
- *Two connected organs are represented by two edges having a common node.*
- *The root is the node below the edge representing the internode at the basis of the plant.*
- *The label of an edge is given by the botanical information of the associated organ: its nature (bud, leaf, internode, ...) and its characteristics (MC, CA, ...).*
- *The label of a node (different from the root) is the label of the edge below it. The root has the label of the edge above it.*

Such PRT is called the labelled PRT associated to the plant P.

N.B.: the labelled PRT defined in Definition 5 is unique for a given plant structure and the converse is true.

Since a plant can be represented by a labelled PRT, it has an associated extended Dyck word. Let $\mathcal{O}^{\mathcal{S}}$ be the minimal set of letters coding for the type of organs needed to describe plant structures generated by $\mathcal{S}$. Generally, the letter b codes for a bud, m for an internode, L for a leaf and F for a fruit. In the same way, let $\mathcal{C}^{\mathcal{S}}$ be the minimal set of all possible characteristics associated to $\mathcal{S}$. In Figure 5, the extended Dyck word associated to the plant (*i.e.* to its PRT) is thus $w = z_{m,1}z_{m,2}z_F z'_F z_{b,2}z'_{b,2}z'_{m,2}z_{m,1}z_{b,2}z'_{b,2}z_{b,1}z'_{b,1}z_L z'_L z'_{m,1}z'_{m,1}$ where the first letter of a label represents the type of an organ and the second its MC. Note that no MC is associated to leaves and fruits.

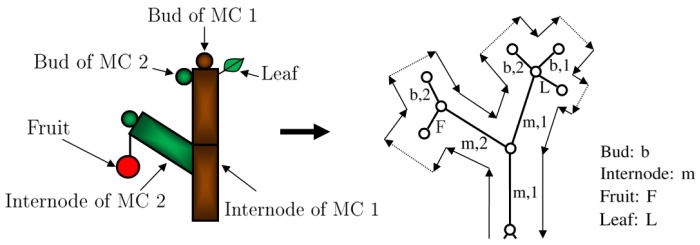

Fig. 5. Correspondence between plants and labelled plane rooted trees. For the sake of clarity, only the labels of edges are represented.

For the sake of clarity, some conventions of coding will be used in the sequel:

- Every organ of nature $o \in \mathcal{O}^{\mathcal{S}}$ having the characteristics $c \in \mathcal{C}^{\mathcal{S}}$ is represented by the symbols o_c and o'_c (instead of $z_{o,c}$ and $z'_{o,c}$).

- For some organs (buds, leaves, fruits, ...), it is not necessary to represent the visit from the child node to the parent node of the associated edge. As a matter of fact, a visit from the parent node to the child node is always immediately followed by a visit from the child node to the parent node. Thus, only the symbol representing the visit from the parent node to the child node is used to describe that type of organ.

By taking into account the previous conventions of coding, the plant of Figure 5 is thus coded by $w = m_1 m_2 F b_2 m_2' m_1 b_2 b_1 L m_1' m_1'$.

In the sequel, the set of all labelled PRTs generated by organogenesis model $\mathcal{S}$ is denoted by $\mathcal{T}^{\mathcal{S}}$. Let $V^{\mathcal{S}} = \{o_c, o_c'\}_{o \in \mathcal{O}^{\mathcal{S}}, c \in \mathcal{C}^{\mathcal{S}}}$ be the minimal alphabet containing all the letters needed to describe plant structures generated by $\mathcal{S}$, $W_{V^{\mathcal{S}}}$ the set of all words built on $V^{\mathcal{S}}$ and $DW_{V^{\mathcal{S}}}$ the set of all extended Dyck words generated by $\mathcal{S}$ (with the previous conventions of notation). Let $D_{V^{\mathcal{S}}} : \mathcal{T}^{\mathcal{S}} \to DW_{V^{\mathcal{S}}}$ be the map which associates for each labelled PRT t its corresponding extended Dyck word $D_{V^{\mathcal{S}}}(t)$. Then, $D_{V^{\mathcal{S}}}$ is a bijection from $\mathcal{T}^{\mathcal{S}}$ to $DW_{V^{\mathcal{S}}}$ (see [12]).

3.3 Growth dynamics and L-systems

In Section 3.2, it was proved that the structure of a plant at a given GC can be represented by an extended Dyck word. We are now interested in the evolution of the structure with respect to its CA starting with a seed (or a bud). As for stochastic organogenesis models driven by multitype branching processes, this evolution is given by a Markovian sequence of random plants indexed by the GCs (see Section 2.2). Thus, it can be described by a sequence $(t^n)_{n \in \mathbb{N}}$ of $\mathcal{T}^{\mathcal{S}}$ (or the corresponding sequence $(D_V(t^n))_{n \in \mathbb{N}}$ of $DW_{V^{\mathcal{S}}}$). To complete the combinatorial framework, we need to define tools which describe the stochastic organogenesis model when the structure is coded by an extended Dyck word (*i.e.* tools which enable the building of $D_V(t^{n+1})$ from $D_V(t^n)$).

Stochastic 0L-systems are well suited to achieve this goal (see [4], [13] and [3]). They are generative parallel rewriting grammars whose production rules are associated to a set of probability distributions.

Definition 6 (Stochastic 0L-system). *Let V be an alphabet and W_V the set of all words built on V. A stochastic 0L-system is a construct $L = \langle \omega_a, \pi \rangle$ where:*

- *$\omega_a \in W_V$ is called the axiom.*
- *π is a transition matrix from V to W_V (i.e., $\forall (u,v) \in V \times W_V$, $0 \leq \pi_{u,v} \leq 1$ and $\sum_{w \in W_V} \pi_{u,w} = 1$) with a finite number of non-zero components.*

A stochastic 0L-system $L = \langle \omega_a, \pi \rangle$ generates a random sequence of words $(w^n)_{n \in \mathbb{N}}$ built on the alphabet V. By definition, the axiom is the word which initiates the sequence generated by L. Then, $w^0 = \omega_a$. We get w^{n+1} by replacing randomly every letter x of w^n by a word y with a probability $\pi_{x,y}$ (note that the evolution of a letter is independent from the evolution of the other letters). By doing so, we create a Markov chain on W_V: $(w^n)_{n \in \mathbb{N}}$. We can now define a more general class of 0L-systems called stochastic F0L-system, extending the classical definition of F0L-system ([14], p. 89) to the stochastic case:

Definition 7 (Stochastic F0L-system). *Let V be an alphabet and W_V the set of all words built on V. A stochastic F0L-system is a construct $L = \langle A, \pi \rangle$ where:*

- *A is a non empty subset of W_V (called the set of axioms of L).*

- *for every $\omega_a \in A$, $L[\omega_a] = \langle \omega_a, \pi \rangle$ is a stochastic 0L-system (called component system of L).*

The stochastic organogenesis model $\mathcal{S}$ can be described by a stochastic F0L-system $L = \langle W_{V^{\mathcal{S}}}, \pi \rangle$. The evolution rules of $\mathcal{S}$ are described by a set of stochastic automata (cf Section 2.2). This set is used to fill the transition matrix π. Let us take the example of Figure 6. From the stochastic automata, we deduce $\pi_{b_1,b_1} = p$, $\pi_{b_1,m_1b_2b_1m_1'} = 1-p$ and $\pi_{b_2,m_2b_2m_2'} = 1$. Note that the evolution rules concerning the internodes of MC 1 and MC 2 are not specified. In that case, it means that they stay in the same state from one GC to another. Therefore, $\pi_{m_1,m_1} = 1$ and $\pi_{m_2,m_2} = 1$. All the other components of π are equal to 0.

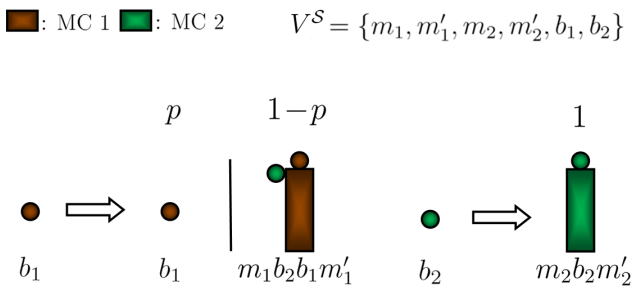

Fig. 6. Stochastic automata and the associated stochastic F0L-system

The stochastic F0L-system $L = \langle W_{V^{\mathcal{S}}}, \pi \rangle$ generates a Markov chain on $DW_{V^{\mathcal{S}}}$, $(w^n)_{n \in \mathbb{N}}$, coding for plant structures. The associated Markov kernel Π (called Markov kernel associated to L) can be built very easily from π (see [10] for more details). For all $(u, v) \in W_{V^{\mathcal{S}}} \times W_{V^{\mathcal{S}}}$, $(\Pi^n)_{u,v}$ is the probability to get the word v by using the stochastic 0L-system $L[u]$ after n steps. w^0 codes for the seed (*i.e.* a bud of MC 1). Since the evolution rules of $\mathcal{S}$ are contained in π, w^{n+1} is built randomly from w^n by using the same rules as those described by $\mathcal{S}$. Therefore, studying the evolution of plant structures generated by $\mathcal{S}$ is completely equivalent to studying the sequence $(w^n)_{n \in \mathbb{N}}$ from a combinatorial point of view.

In the sequel, $L = \langle W_{V^{\mathcal{S}}}, \pi \rangle$ will denote a stochastic F0L-system associated to the stochastic organogenesis model $\mathcal{S}$. When no confusion arises concerning the model used, the letter $\mathcal{S}$ will not be specified in the corresponding sets. In that case, $L = \langle W_V, \pi \rangle$. Π will denote the Markov kernel associated to L. For all $s \in W_V$, $DW_V^{\pi,n}[s]$ $(\subset DW_V)$ denotes the set of all possible extended Dyck words generated by the component system $L[s]$ after n steps. Let $\mathcal{T}^{\pi,n}[s] = D_V^{-1}(DW_V^{\pi,n}[s])$ be the set of all labelled PRTs $(\mathcal{T}^{\pi,n}[s] \subset \mathcal{T})$ generated by $L[s]$ after n steps.

4 The symbolic method

The symbolic method is a very effective method to analyse combinatorial structures and, as a consequence, plays an important role in analytic combinatorics (see [5] for more details). As far as plants are concerned, it enables us to compute not only the distribution of the number of organs (of any type) but also of specific structures in the plant architecture.

4.1 Combinatorial class and Generating function

In this section, basic tools of combinatorics are recalled (see [5] for more details).

Definition 8 (Combinatorial Class). *A combinatorial class, or simply a class, is a finite or denumerable set on which a size function is defined, satisfying the following conditions:*

- *the size of an element is a non-negative integer.*
- *the number of elements of any given size*

is finite.

For example, for all $n \in \mathbb{N}$, $DW_V^{\pi,n}[s]$ is a combinatorial class. Many size functions can be defined (counting the number of letters coding for internodes, ...).

Definition 9 (Stochastic Combinatorial Class). *A stochastic combinatorial class is a set $SC = \{(t, p_t), t \in C\}$ such that:*

- *C is a combinatorial class.*
- *$\forall t \in C$, $0 \leq p_t \leq 1$.*
- *$\sum_{t \in C} p_t = 1$.*

The set $SDW_V^{\pi,n}[s] = \{(w, (\Pi^n)_{s,w}), w \in DW_V^{\pi,n}[s]\}$ is a stochastic combinatorial class.

Definition 10 (Generating Function (= GF) associated to a size function in a stochastic combinatorial class). *Let C be a combinatorial class and $SC = \{(t, p_t), t \in C\}$ a stochastic combinatorial class. Let m be a size function in C. The generating function Ψ associated to m in SC is a mapping from $[0, 1]$ to $[0, 1]$ defined as follows:*

$$\forall z \in [0,1], \quad \Psi(z) = \sum_{t \in C} p_t z^{m(t)}$$

GF are very useful to analyse a stochastic combinatorial class SC from a probabilistic point of view. Suppose we are interested in getting the distribution of a particular characteristic c in SC. Let m be the size function ($SC \to \mathbb{N}$) such that, for all $t \in SC$, $m(t)$ gives the number of c in the element t. By reordering the terms of the GF,we get the following power series:

$$\forall z \in [0,1], \quad \psi(z) = \sum_{t \in C} p_t z^{m(t)} = \sum_{k \in \mathbb{N}} p(k) z^k$$

$p(k)$ is the probability to get k characteristics c in an element of SC. Therefore, by determining the coefficients of the GF associated to s in SC, we get the distribution of c in SC.

4.2 Description of the method

Suppose we are interested in getting the distribution of a particular structure (a specific sequence of phytomers, a particular element in the plant, ...) in a plant of CA n generated by the stochastic organogenesis model $\mathcal{S}$. This particular structure can be coded on the alphabet V by a word u. Therefore, this problem is equivalent to determining the distribution of a pattern u in the stochastic combinatorial class $SDW_V^{\pi,n}[s]$.

Definition 11 (Counting Function). *The counting function c is a map from $W_V \times W_V$ to $\mathbb{N}$ such that, for all $(w, u) \in W_V \times W_V$, $c(w, u)$ gives the number of patterns u in the word w.*

For all $u \in W_V$ and $n \in \mathbb{N}$, the mapping $w \mapsto c(w, u)$, from $DW_V^{\pi,n}[s]$ to $\mathbb{N}$, is a size function for the combinatorial class $DW_V^{\pi,n}[s]$. Therefore, to get the distribution of a pattern u in $SDW_V^{\pi,n}[s]$, we need to compute the GF associated to the size function $w \mapsto c(w, u)$ in $SDW_V^{\pi,n}[s]$. In the sequel, we will call GF associated to a pattern u in $SDW_V^{\pi,n}[s]$ the previous GF. It will be denoted by $\psi^n[s](z)$:

$$\forall z \in [0,1], \ \psi^n[s](z) = \sum_{w \in DW_V^{\pi,n}[s]} (\Pi^n)_{s,w} z^{c(w,u)} = \sum_{k \in \mathbb{N}} P^{n,s}(k) z^k$$

where $P^{n,s}(k)$ is the probability to get k patterns u in a plant structure generated by $L[s] = \langle s, \pi \rangle$ after n steps. However, this GF is rarely determined directly. Usually, we get it from functional equations which are most of the time recurrence relations between $\psi^n[s]$ and $\psi^{n+1}[s]$. To obtain these equations, we use a symbolic approach as developed in [5].

Let C be a combinatorial class and $SC = \{(t, p_t), t \in C\}$ a stochastic combinatorial class. Suppose we are interested in a particular size function m taking its argument in $\mathcal{C}$. Thus, we

want to determine the GF ψ associated to m in $\mathcal{SC}$. The idea of the symbolic method is to build an equation which decomposes SC into smaller classes either of the same type or of simpler types (in the sequel, we will refer to this equation as the set equation). Then, it is transformed into an equation composed of the GFs of the previous combinatorial classes. When it is possible, we can solve directly the transformed equation and we get Ψ. When the structure of the class SC is too complex, we extract recurrence relations satisfied by the coefficients of Ψ from the transformed equation.

When dealing with plants, for a given $n \in \mathbb{N}$, we have $C = DW_V^{\pi,n}[s]$ and $SC = SDW_V^{\pi,n}[s]$ where s is an extended Dyck word coding for a seed. The idea is to decompose $DW_V^{\pi^n}[s]$ into a combination of sets of type $DW_V^{\pi^k}[s']$ with $k < n$ and $s' \in W_V$. By doing so, we get a set equation which can be transformed into an equation composed of GFs.

The symbolic method can be decomposed into the following steps:

1. Identify the structure of the plant and all stochastic behaviours.

2. Deduce the associated stochastic automata.

3. Find the appropriate alphabet V to describe the plant and write the stochastic F0L-system $L = \langle W_V, \pi \rangle$.

4. Identify the pattern $u \in W_V$ whose distribution you want to compute.

5. Write the structural property associated to each $DW_V^{\pi,n}[s]$ for $n \in \mathbb{N}$ where s is an extended Dyck word coding for a seed and try to decompose it into a combination of sets of type $DW_V^{\pi^k}[s']$ with $k < n$ and $s' \in W_V$.

6. Write the transformed equation satisfied by the GF $\psi^n[s]$ associated to u in $SDW_V^{\pi,n}[s]$ for each $n \in \mathbb{N}$.

7. Either solve directly the set of transformed equations or find a recurrence relation between $\psi^n[s]$ and $\psi^{n+1}[s]$ with $n \in \mathbb{N}$.

8. Extract from these equations the coefficients of the GFs.

4.3 Examples

4.3.1 Example with simple elements

We want to compute the distribution associated to the number of internodes in the following plant growth model: a plant with only one MC and a dormancy probability p. The behaviour of a bud is characterized by Figure 7. The alphabet is simply $V = \{m, m', b\}$ where m codes for an internode and b for a bud.

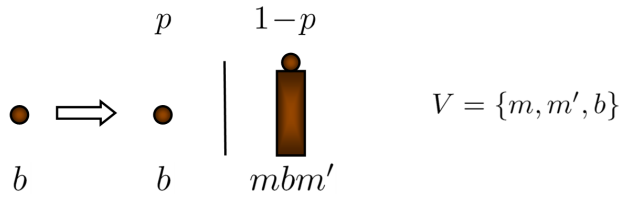

Fig. 7. Plant with one morphological category and dormancy probabilities

The transition matrix of the associated stochastic F0L-system $L = \langle W_V, \pi \rangle$ can be easily deduced from Figure 7. We have $\pi_{b,b} = p$, $\pi_{b,mbm'} = 1 - p$, $\pi_{m,m} = 1$ and $\pi_{m',m'} = 1$. All the other components of π are equal to zero. The pattern of interest is m. From Figure 7, we deduce that a word $w \in DW_V^{\pi,n+1}[b]$ can be:

- either a word $v \in DW_V^{\pi,n}[b]$ with a probability p,

- or a word of the form mvm' with $v \in DW_V^{\pi,n}[b]$ with a probability $1 - p$.

Let $\psi^n[b]$ be the GF associated to m in $DW_V^{\pi,n}[b]$:

$$\psi^n[b](z) = \sum_{w \in DW_V^{\pi^n}[b]} (\Pi^n)_{b,w} \quad z^{c(w,m)} = \sum_{k \in \mathbb{N}} P^{n,b}(k) z^k$$

where $P^{n,b}(k)$ is the probability to get k internodes in a plant structure generated by the organogenesis model described in Figure 7 after n GCs. Then, we have:

$$\psi^{n+1}[b](z) = \sum_{v \in DW_V^{\pi^n}[b]} p(\Pi^n)_{b,v} \quad z^{c(v,m)} + \sum_{v \in DW_V^{\pi^n}[b]} (1-p)(\Pi^n)_{b,v} \quad z^{c(mvm',m)}$$

Given that $c(mvm', m) = c(v, m)+1$, we have:

$$\psi^{n+1}[b](z) = p \sum_{v \in DW_V^{\pi^n}[b]} (\Pi^n)_{b,v} \quad z^{c(v,m)} + (1-p)z \sum_{v \in DW_V^{\pi^n}[b]} (\Pi^n)_{b,v} \quad z^{c(v,m)}$$

and then:

$$\psi^{n+1}[b](z) = p\psi^n[b](z) + (1-p)z\psi^n[b](z) = (p + (1-p)z)\psi^n[b](z) \quad (1)$$

Given that $\psi^0[b] = 1$, the solution of the previous geometric progression is then:

$$\psi^n[b](z) = (p + (1-p)z)^n$$

We deduce $P^{n,b}(k) = \binom{n}{k} p^{n-k}(1-p)^k$ if $0 \leq k \leq n$ and 0 otherwise.

N.B.: as detailed in [10], the underlying stochastic process in this section is that of a Galton-Watson multitype branching process (see [8] and [9]). As a matter of fact, let B_n and M_n be two random variables on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ (where $\mathbf{P}$ is a probability measure) such that B_n and M_n give respectively the number of letters b and m in a word generated randomly by either the component system $L[b]$ or the component system $L[m]$ after n steps. Then, the sequence of random vectors $\Big((B_n, M_n)\Big)_{n \in \mathbb{N}}$ is a Galton-Watson multitype branching process. Let Φ_n be the probability generating function associated to (B_n, M_n) for all $n \in \mathbb{N}$. In that case, Equation 1 is simply the translation of the composition formula for branching processes: $\Phi_{n+1} = \Phi_1(\Phi_n)$ ([15]).

4.3.2 Example with a complex structure

In this section, we introduce an example which illustrates well the benefit of the symbolic approach. We want to compute the distribution associated to the number of 'Y-structures' (see Figure 8).

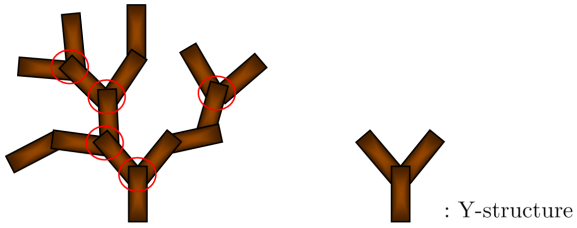

Fig. 8. Example of Y-structure

The plant development model is given by Figure 9.

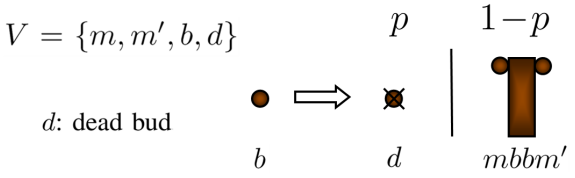

Fig. 9. Leeuwenberg type growth model with death probability

The alphabet is $V = \{m, m', b, d\}$ where d represents a dead bud and the stochastic F0L-system can be easily deduced from Figure 9. We have $\pi_{b,d} = p$, $\pi_{b,mbbm'=1-p}$, $\pi_{m,m} = 1$, $\pi_{m',m'} = 1$ and $\pi_{d,d} = 1$. All the other components of π are equal to zero. Counting the number of 'Y-structures' is equivalent to counting the number of patterns $m'm$. Thus, the pattern of interest is $m'm$. As for the structural property of $DW_V^{\pi,n}[s]$, we have to break down

the structure of a PRT in a way which highlights how Y-structures appear in the topology and how they are connected to substructures (see Figure 9). From Figure 9, we deduce that a word $w \in DW_V^{\pi,n+1}[b]$ can be:

- either a dead bud d with a probability p,
- or a word of the form mvm' with $v \in DW_V^{\pi,n}[bb]$ with a probability $1-p$. In that case, v represents a branched structure.

Let $\psi^n[b]$ be the GF associated to $m'm$ in $DW_V^{\pi,n}[b]$:

$$\psi^n[b](z) = \sum_{w \in DW_V^{\pi^n}[b]} (\Pi^n)_{b,w} \quad z^{c(w,m'm)} = \sum_{k \in \mathbb{N}} P^{n,b}(k) z^k$$

where $P^{n,b}(k)$ is the probability to get k Y-structures in a plant generated by the organogenesis model described in Figure 9 after n GCs. Then, we have:

$$\begin{aligned}\psi^{n+1}[b](z) &= pz^{c(d,m'm)} + \sum_{v \in DW_V^{\pi^n}[bb]} (1-p)(\Pi^n)_{bb,v} z^{c(mvm',m'm)} \\ &= p + (1-p)\psi^n[bb](z) \end{aligned} \tag{2}$$

Now, we need to find a recurrence relation for $\psi^n[bb](z)$. It can be deduced straightforward by the decomposition of Figure 10:

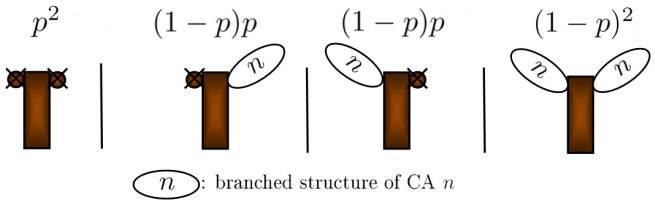

Fig. 10. Decomposition of branched structures of CA $n+1$

We deduce that a word $w \in DW_V^{\pi,n+1}[bb]$ (*i.e.* a branched structure) can be:

- two dead buds dd with a probability p^2.
- one branched structure mvm' with $v \in DW_V^{\pi,n}[bb]$ and one dead bud d (either vd or dv) with a probability $p(1-p)$ for each case.
- two branched structures mv_1m' and mv_2m' with $(v_1, v_2) \in DW_V^{\pi,n}[bb] \times DW_V^{\pi,n}[bb]$ with a probability $(1-p)^2$.

Therefore:

$$\begin{aligned}\psi^{n+1}[bb](z) &= p^2 z^{c(dd,m'm)} \\ &+ \sum_{v \in DW_V^{\pi^n}[bb]} (1-p)p(\Pi^n)_{bb,v} z^{c(mvm'd,m'm)} \\ &+ \sum_{v \in DW_V^{\pi^n}[bb]} (1-p)p(\Pi^n)_{bb,v} z^{c(dmvm',m'm)} \\ &+ \sum_{(v_1,v_2) \in \left(DW_V^{\pi^n}[bb]\right)^2} (1-p)^2 (\Pi^n)^2_{bb,v} z^{c(mvm'mvm',m'm)}\end{aligned}$$

Since $c(mvm'd, m'm) = c(dmvm', m'm) = c(v, m'm)$ and $c(mvm'mvm', m'm) = 2c(v, m'm) + 1$, we get:

$$\psi^{n+1}[bb](z) = p^2 + 2(1-p)p\psi^n[bb](z) + (1-p)^2 z \left(\psi^n[bb](z)\right)^2 \tag{3}$$

By identifying the coefficients of the power series involved in Equations 2 and 3 , we get the distribution of Y-structures. The same method would work to compute, for example, the distributions associated to the number of apexes (*i.e.* terminal nodes in a tree).

5 Application to parameter identification

In plant stochastic organogenesis model, the parameters identification of the automata may be delicate and necessitates an important sampling work. For stochastic organogenesis models driven by a multitype branching process, classical methods based on branching processes have already been established (see [16] and [17]). These methods rely on the calibration of the parameters so that the theoretical mean and variance associated to the number of phytomers (computed from the model) are the closest to the experimental ones. However, phytomers are not always easily identifiable in

a plant. Therefore, the idea is to use a botanical structure that is more easy to identify and to count (for example apices or Y-structures, cf Section 4.3.2). By confronting the theoretical distribution of this structure to the experimental one, we are able to find the best set of parameters that will give to the model the closest behaviour to real plants. However, this theoretical distribution can rarely be determined with classical branching process methods. In that case, the symbolic method of Section 4 provides a good alternative.

Let $\mathcal{S}$ be a stochastic organogenesis model driven by a multitype branching process and $L = \langle W_V, \pi \rangle$ the associated stochastic F0L-system. Suppose we are interested in computing the theoretical distribution of a particular structure in random plant architectures generated by $\mathcal{S}$. This particular structure is coded on the alphabet V by the word u. Let $\psi^n[s]$ be the GF associated to u in $SDW_V^{\pi,n}[s]$ where $s \in W_V$ is the word coding for a seed. As mentioned in Section 4.2, the theoretical distribution is given by the coefficient of $\psi^n[s]$ seen as power series:

$$\forall z \in [0,1],\ \psi^n[s](z) = \sum_{w \in DW_V^{\pi,n}[s]} (\Pi^n)_{s,w} z^{c(w,u)} = \sum_{k \in \mathbb{N}} p^{n,s}(k) z^k$$

Since $card(DW_V^{\pi,n}[s]) < \infty$ (the stochastic automata can only generate a finite number of structures), then, for all $n \in \mathbb{N}$, $\max\{c(w,u) | w \in DW_V^{\pi,n}[s]\}$ exists and is finite. In that case, we set:

$$\forall n \in \mathbb{N},\quad L^n = \max\{c(w,u) | w \in DW_V^{\pi,n}[s]\}$$

Thus, for all $l > L^n$, $p^{n,s}(l) = 0$. Let ϕ^n be a vector in $[0,1]^{L^n+1}$ such that:

$$\forall n \in \mathbb{N},\ \phi^n = (p^{n,s}(0), p^{n,s}(1), \ldots, p^{n,s}(L^n))$$

Generally, the symbolic method leads to a recursive equation between $\psi^n[s]$ and $\psi^k[r]$ with $k \in K$ and $r \in R$ where K and R are respectively finite subsets of $\{0, \ldots, n-1\}$ and W_V. For the sake of clarity, we will suppose that the method gives us a recursive equation between $\psi^n[s]$ and $\psi^{n-1}[s]$ (the extension to the general case is straightforward). Therefore, by identifying the coefficients of the power series, we get a set of recurrence relations between the components of ϕ^n and ϕ^{n-1} which enables the building of ϕ^n from ϕ^{n-1}. The stochastic automata depend on a set P of parameters which have a botanical meaning (survival probability,...). As a consequence, the set of recurrence relations between the components of ϕ^n and ϕ^{n-1} depends also on P and, thus, $\phi^n = \phi^n(P)$.

Suppose we have a plot of plants of CA $N \in \mathbb{N} \setminus \{0\}$. For each of these plants, we measure the number of characteristics u. By doing so, we get the experimental distribution of the particular structure: $\{p^{exp}(l)\}_{l \in \mathbb{N}}$. Suppose that there exists $l > L^N$ such that $p^{exp}(l) \neq 0$, then the organogenesis model is not well defined. As a matter of fact, for $l > L^N$, $p^{N,s}(l) = 0$. In that case, ϕ^N cannot get as close as we want to the experimental distribution. In that case, the organogenesis model (*i.e.* the automata) needs to be modified. Then, the model $\mathcal{S}$ is said to be well defined if:

$$\min\{l \in \mathbb{N} | P^{exp}(l) = 0\} > L^N$$

Let ϕ^{exp} be a vector in $[0,1]^{L^N+1}$ such that:

$$\phi^{exp} = (p^{exp}(0), p^{exp}(1), \ldots, p^{exp}(L^N))$$

The set of parameters P is estimated by the least square estimator $\hat{P}$:

$$\hat{P} = \underset{P \in [0,1]^{card(P)}}{\operatorname{argmin}} \left\| \phi^{exp} - \phi^N(P) \right\|_2^2$$

where, for all $\alpha = (\alpha_0, \ldots, \alpha_{L^N}) \in [0,1]^{L^N+1}$, $\|\alpha\|_2 = \left(\sum_{i=0}^{L^N} (\alpha_i)^2\right)^{1/2}$.

Several optimisation algorithms can be used to find $\hat{P}$. One of the most appropriate algorithms to solve the minimisation problem is the Levenberg-Marquardt algorithm (see [18]).

6 Conclusion

In this article, a symbolic method was set to analyse complex structures in plants whose organogenesis is driven by a multitype branching process. To that purpose, a new combinatorial framework was introduced. Plant structure is represented by a plane rooted tree and, as a consequence, can be coded by a Dyck word. The evolution rules of the organogenesis model are given by a stochastic 0L-system. By doing so, the evolution of plant structure is coded by a Markovian sequence of Dyck words. Therefore, studying plant structure and its development is completely equivalent to studying the Markovian sequence from a combinatorial point of view. A symbolic method was then established and enables the computation of the distribution associated to the number of complex structures in plant topology.

This result has numerous applications. For instance, in this article, we have shown that such a method can be used for parameter identification. Moreover, it can also be used to compare stochastic organogenesis models. As a matter of fact, by confronting the theoretical distribution of a structure of a given type, we are able to choose the model which has the closest behaviour to that of real plants.

References

[1] A. Lindenmayer, "Mathematical models for cellular interactions in development. i. filaments with one-sided inputs," *Journal of Theoretical Biology*, vol. 18, pp. 280–289, 1968.

[2] A. Smith, "Plants, fractals and formal languages," *Computer Graphics (SIGGRAPH 84 Conference Proceedings)*, vol. 18, no. 3, pp. 1–10, 1984.

[3] W. Kurth, *Growth grammar interpreter GROGRA 2.4: A software tool for the 3-dimensional interpretation of stochastic, sensitive growth grammars in the context of plant modelling. Introduction and Reference Manual.* Berichte des Forschungszentrums Waldokosysteme der Universitat Gottingen, Ser. B, Vol. 38, 1994.

[4] P. Prusinkiewicz and A. Lindenmayer, *The Algorithmic Beauty of Plants.* Springer-Verlag, New-York, 1990.

[5] P. Flajolet and R. Sedgewick, *Analytic Combinatorics.* Cambridge University Press, 2009.

[6] D. Barthélémy and Y. Caraglio, "Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny," *Annals of Botany*, vol. 99, no. 3, pp. 375–407, 2007.

[7] M.-Z. Kang, P.-H. Cournède, P. de Reffye, D. Auclair, and B.-G. Hu, "Analytical study of a stochastic plant growth model: application to the GreenLab model," *Mathematics and Computers in Simulation*, vol. 78, no. 1, pp. 57–75, 2008.

[8] C. Mode, *Multitype branching processes: Theory and applications.* American Elsevier Publishing Co. Inc, New York, 1971.

[9] K. Athreya and P. Ney, *Branching Processes.* Dover Publications, 2004.

[10] C. Loi and P.-H. Cournède, "Generating Functions of Stochastic L-Systems and Application to Models of Plant Development," *Discrete Mathematics and Theoretical Computer Science Proceedings*, vol. AI, pp. 325–338, 2008.

[11] J. Riordan, *An Introduction to Combinatorial Analysis.* Courier Dover Publications, 2002.

[12] D. Knuth, *The Art of Computer Programming.* Addison-Wesley Professional, 1997 (3rd edition), vol. 1.

[13] J. Françon, "Sur la modélisation informatique de l'architecture et du développement des végétaux." in *2ème Colloque International: L'Arbre. Institut de Botanique, Montpellier, France*, 1990.

[14] G. Rozenberg and A. Salomaa, *The Mathematical Theory of L-systems.* Academic Press, New York, 1980.

[15] T. Harris, *The theory of branching processes.* Springer, Berlin, 1963.

[16] Y. Guédon, D. Barthélémy, Y. Caraglio, and E. Costes, "Pattern analysis in branching and axillary flowering sequences," *Journal of Theoretical Biology*, vol. 212, pp. 481–520, 2001.

[17] M. Kang, P.-H. Cournède, J.-P. Quadrat, and P. de Reffye, "A stochastic language for plant topology," in *Plant growth Modeling, simulation, visualization and their Applications.*, T. Fourcaud and X. Zhang, Eds. IEEE Computer Society (Los Alamitos, California), 2007.

[18] J. Moré, *The Levenberg-Marquardt Algorithm: Implementation and Theory*, Springer Berlin / Heidelberg ed. Springer, 2006, pp. 105–116, ISBN: 978-3-540-08538-6.

Cédric Loi graduated from the French engineer school Ecole Centrale Paris in 2007 where he studied applied mathematics. Then, he received a master degree in statistics from the university Paris XI. He is currently a PhD student in the laboratory of applied mathematics at Ecole Centrale Paris. His thesis is about the study of plant organogenesis models driven by multitype branching processes and their parametric identification.

Paul-Henry Cournède graduated from Ecole Centrale Paris and the University of Cambridge in 1997, and received his PhD in Applied Mathematics from Ecole Centrale Paris in 2001. He is currently full professor at the department of Mathematics of Ecole Centrale Paris. His research activity in the laboratory of Applied Mathematics and Systems focuses on the mathematical modeling of plant growth. He belongs to the research group Digiplante (INRIA, Ecole Centrale Paris, CIRAD).

Jean Françon is a retired professor of the University of Strasbourg. In the 70's and 80's his research at the research institute for advanced mathematics (IRMA) and at the centre for nuclear studies (CEN) in Strasbourg led to pioneering works in combinatorics and in the analysis of algorithms. He then participated to the creation of the "Equipe d'Informatique Géométrique et Graphique" in the laboratory of Image, Computer and Remote Sensing Sciences at the University of Strasbourg (Louis Pasteur University) and also started working on geometrical modelling and discrete geometry. His interest for the potential applications of combinatorics to models of plant growth dates back to the 80's but has recently found new echoes.